SORT 37 (2) July-December 2013, 153-174

A nonparametric visual test of mixed hazard models

Jaap Spreeuw[1], Jens Perch Nielsen[2] and Søren Fiig Jarner[3]

Abstract

We consider mixed hazard models and introduce a new visual inspection technique capable of detecting the credibility of our model assumptions. Our technique is based on a transformed data approach, where the density of the transformed data should be close to the uniform distribution when our model assumptions are correct. To estimate the density on the transformed axis we take advantage of a recently defined local linear density estimator based on filtered data. We apply the method to national mortality data and show that it is capable of detecting signs of heterogeneity even in small data sets with substantial variability in observed death rates.

MSC: 62F10, 62N01, 62N02, 62P05.

Keywords: Mortality data, frailty models, visual inspection.

1. Introduction

There is an increasing use of mortality models to answer a number of pension related questions. Mortality tables and their estimation have always been of importance while calculating appropriate prices of risk products depending on individuals' survival. More recently, mortality models are being used in more complex models assessing the value of financial products incorporating survival in a variety of ways. Financial users of mortality models are therefore not only actuaries nowadays, but also investors looking for opportunities in survival bonds and other packages of survival risks. Different purposes of mortality models lead to different measures of quality.

[1] Faculty of Actuarial Science and Insurance, Cass Business School, City University London, 106 Bunhill Row, London, EC1Y 8TZ, UK. E-mail: j.spreeuw@city.ac.uk

[2] Faculty of Actuarial Science and Insurance, Cass Business School, City University London, 106 Bunhill Row, London, EC1Y 8TZ, UK.

[3] Danish Labour Market Supplementary Pension Fund, Kongens Vænge 8, 3400 Hillerød, Denmark.

Received: October 2012
Accepted: July 2013

In this paper we develop a visualization technique that seems useful for the individual assessment of the quality of a mortality model. One application we are thinking of is forecasting of mortalities that is a basic building block for the financial pricing of survival, but also a useful tool in asset liability management of pension portfolios. Typically, relatively simple parametric mortality models including calendar effects are used as starting point for mortality forecasts. The calendar effect is the explicit tool for the forecast and is often isolated and estimated through standard time series methodology. A perfect historical fit of the past is therefore not always what the mortality modeller is looking for. Often it is more important to have an overall good fit, without too systematic deviations giving reliable and meaningful forecasts. These latter objectives are not easy to generalize to some quantitative model that can be tested. Often simple mortality models are rejected, simply because mortality data often is nationwide and sufficiently abundant to inform relatively complex underlying parametric structures. Therefore, a test rejecting our simple model is often not what we want. We do know that our simple model is not accurate, we do not want an excessive fit, what we want is a good, intuitive and reliable forecast.

When modelling mortality of a population, there is a variety of potentially suitable lifetime data models available. Potential models differ in levels of complexity and they try to capture different features of data. Specific parametric life tables combined with time series forecasts are omnipresent in the actuarial and demographic literature.

The literature about parametric mortality projection has been developing rapidly in the last few years. Recent reviews of mainstream mortality forecasting models can be found in Cairns et al. (2009), Cairns et al. (2011), Dowd et al. (2010a,b) and Haberman and Renshaw (2011). Cairns et al. (2009) compare eight models on the basis of several desirable ex post qualitative properties (like model parsimony, transparency, possibility to generate sample paths, presence (or absence) of cohort effects and ability to achieve a nontrivial correlation structure) and quantitative criteria (consistency with historical data and robustness of parameter estimates). Six of these models are subject of subsequent investigation by Dowd et al. (2010a,b) and Cairns et al. (2011). These include the original Lee-Carter model (Lee and Carter, 1992), the basic age-period-cohort model by Renshaw and Haberman (2006), an alternative age-period-cohort model by Currie (2006), the original Cairns-Blake-Dowd model as launched in Cairns et al. (2006), and two extensions thereof. The six models are the subject of formal goodness-of-fit tests in Dowd et al. (2010a) and backtesting in Dowd et al. (2010b). Cairns et al. (2011) judges these models on the basis of ex ante qualitative aspects like biological reasonableness, plausibility of forecast levels of uncertainty in projections at several ages, and robustness of forecasts. In all these papers, the mortality data applied was confined to those of individuals aged 60 or above. Haberman and Renshaw (2011), concentrating on the key factors of life expectancy and annuity values, first conduct a detailed comparison of the several models at pensioner ages. Apart from the models in the above papers, they also consider special cases of the Renshaw and Haberman (2006) model in their study. Later on, they extend the age range and involve the model by Plat (2009) and several variants thereof.

The stability of the forecast depends crucially on the choice of the parametric form. Generally, a complex model with many parameters is not a good choice even though such models might be selected from classical mathematical statistical model selection designed for in-sample prediction. Models with many parameters generally fit data better than models with fewer parameters. On the other hand, a large number of parameters are harder to forecast than fewer parameters. Forecasting uncertainty increases dramatically with the number of parameters. Thus, to obtain reliable forecasts we want models which describe the key features of data with as few parameters as possible.

The purpose of this paper is to introduce a visual diagnostic tool which can be used to guide us when choosing a parametric model. A good parametric model is a simple model without obvious systematic errors. That model could be chosen by the well informed statistician working with the particular mortality forecast application in mind. Our visual diagnostic tool will be just one helpful tool in the overall mathematical statistical toolbox. Our method is inspired from recent developments in extreme value estimation, where transformations of data give visual information on the quality of the distributional fit in the tail. This recent methodology has found its way into insurance pricing and also the related field of operational risk. For a comprehensive overview of this new transformation methodology in the latter context, see Bolancé et al. (2012a).

The transformation based method can for example compare the performance of several candidate models for a data set at hand. Assume we were told by an oracle what the exact true distribution is, then we would transform our data using this oracle information such that our transformed data would exactly originate from a uniform distribution. Now we do not have access to any oracles. However, if we take some estimated parametrically fitted survival distribution as defining our transformation, then any detectable deviance on the transformed scale from the uniform distribution implies deviances of the parametric distribution used in the transformation step from the underlying true distribution. Our methodology uses a nonparametric smooth kernel estimator on the transformed scale. One difficulty we meet here is that our data is classical survival data that is not independent identically distributed. We therefore use a recent local linear kernel density estimator – specifically the one of Nielsen et al. (2009) – that is adjusted for the truncation and censoring pattern we meet in our data. Comparison between different underlying suggested parametric models are carried out by first estimating these parametric models and then to investigate through visual inspection, whether the density of the transformed data indeed looks uniform.

If the underlying parametric model under investigation would be true, the estimated density should be close to one over the unit interval. Therefore different underlying parametric models can be visualized and compared on the transformed scale. In principle, the densities could also be estimated and compared on the original scale. However, there are several visual and estimational advantages to working on the transformed scale. One of these is that our method makes maximal use of sparse and volatile data and is thus particularly well suited to explore how potential models describe the mortality at ad-

vanced ages where exposure is invariably limited. We test our method using data from nations of different size: USA, United Kingdom, Denmark and Iceland.

Although the main focus of our paper is to model human mortality, it is worthwhile mentioning that our methodology is applicable to any probability density model, whether it concerns human survival or not.

1.1. Mixed hazard models

Frailty theory offers a possible explanation to the presence of an old-age mortality plateau. According to this theory populations are heterogeneous with some people being more frail, i.e. having a higher hazard rate, than other people. Since persons with high hazard rates tend to die sooner than persons with low hazard rates old age groups will be dominated by low frailty persons and this effect reduces the rate of increase at the population level.

Frailty models were introduced in the demographic literature by Vaupel et al. (1979). In a multiplicative frailty model, an individual's hazard rate consists of two parts, namely a certain standard intensity and a certain nonnegative random variable, the frailty, acting multiplicatively on the standard intensity. A Gompertz or Makeham specification is usually taken for the standard intensity, although sometimes a Weibull model can be seen. Frailty is usually assumed to follow a Gamma distribution, which is known to be mathematically very tractable.

A few publications about frailty modelling appeared in the actuarial literature. Wang and Brown (1998) use the Gompertz-Gamma or Perks model to graduate mortality improvement factors in a Society of Actuaries' Life Table. Butt and Haberman (2004) employ Generalized Linear Models to graduate mortality of insured lives. They consider three mixture models, namely i) Perks; ii) modified Perks, and iii) Gompertz-Inverse Gaussian. The authors conclude that the Perks model fits the data best. An overview of heterogeneity models in life insurance is given in Olivieri (2006), while Jones (1998) develops a multiple state model to measure the impact of frailty on the propensity to lapse a policy. Finally, Li et al. (2009) extend the Lee-Carter model by allowing for unobserved heterogeneity within a cell, determined by age and time.

In this paper we illustrate our methodology in the one dimensional case. Most forecasting models operate with a multiplicative relationship between age effect and time effect. To visualize the fit of the age effect, one would then have to divide out the estimated time effect and vice versa to visualize the time effect only.

We are happy to say that our paper – diffused in preliminary versions – already has inspired a number of other works in mathematical and computational statistics. It has for example been cited in the three recent papers Gámiz-Pérez et al. (2013a,b,c).

1.2. Outline

The set-up of this paper is as follows. In Section 2 we present the visual inspection technique in detail. Both the continuous-time framework with transformed counting processes and the implementation with discrete data is discussed. Section 3 discusses frailty models in general and introduces the class of models we will be using. Section 4 presents the numerical application. For four countries varying significantly in size (United States, United Kingdom, Denmark and Iceland), one data set per country (female period 2006 from the Human Mortality Database) and three different frailty specifications, namely Gamma, Inverse Gaussian, and degenerate (no frailty), we show the estimates as well as the visual inspection technique. In particular, we give a thorough analysis of the mortality at advanced ages that can be extracted from the continuous graphs. Section 5 sets out a conclusion.

2. Visual inspection technique

2.1. Sampling scheme of the survival data

Consider a data set with mortality statistics of n lives. Let for each of these n individuals Y_i be an exposure process with value one when the i'th individual is alive and under observation and let N_i be a counting process taking the value one if the i'th individual has died while under observation. Both Y_i and N_i are functions of the age x. Formally, we assume that N_i is a one-dimensional counting process with respect to an increasing right continuous complete filtration $\mathscr{F}_x,\ x \in \mathscr{R}_+$, i.e. one that obeys *les conditions habituelles*, see Andersen et al. (1993, p. 60). We model the intensity as

$$\lambda_i^c(x) = \mu_\theta(x)\, Y_i(x),$$

where θ belongs to the parameter space Θ of the parameters determining the exact mortality and frailty. The estimator $\widehat{\theta}$ of θ is derived from minimizing the log likelihood of Borgan (1984):

$$l(\theta) = \sum_{i=1}^{n} \int \log\{\mu_\theta(x)\}\, dN_i(x) - \sum_{i=1}^{n} \int \mu_\theta(x)\, Y_i(x) dx,$$

that is maximized over the parameter space Θ.

2.2. Visual inspection by transformations

Assume that some oracle has given us the true underlying c.d.f. F_θ. Then consider the transformed counting processes $\overline{N}_i = N_i \circ F_\theta^{-1}$ defined on $[0,1]$. If our oracle really had

told us the truth, then $\overline{N}_i$ would have stochastic intensity

$$\lambda_i(y) = \alpha(y)\overline{Y}_i(y),$$

where $\overline{Y}_i(y) = \left\{Y_i\left(F_\theta^{-1}(y)\right)\right\}$ with $\alpha(y) = 1/(1-y)$ corresponding to the hazard of the uniform distribution with density

$$f(y) = \alpha(y)\exp\left(\int_0^y -\alpha(s)ds\right) = 1,$$

for $y \in [0,1]$.

Another more statistical term for oracle information is prior information. It is that type of information that is external to the data set at hand. In our application below our prior information will always be some parametric specification of the model and our oracle candidate for the true c.d.f will be $F_{\widehat{\theta}}$, where $\widehat{\theta}$ is the estimated parameter in the specified parametric model. If $F_{\widehat{\theta}}$ really is a good description of the true c.d.f. F, then our data should be uniformly distributed after a transformation by $F_{\widehat{\theta}}$.

To be able to inspect the credibility of our oracle information or prior information or parametric assumptions, we estimate the density f based on the filtered survival data $\left(\overline{N}_1,\overline{Y}_1\right),\ldots,\left(\overline{N}_n,\overline{Y}_n\right)$ on $[0,1]$ and see whether it looks flat. This density estimator should have good boundary correction because it is defined on the transformed axis $[0,1]$. We suggest to use the natural weighted local linear density estimator of Nielsen et al. (2009):

$$\widehat{f}(y) = \sum_{i=1}^{n}\int \overline{K}_{y,b}(y-s)\overline{Y}_i(s)\widehat{S}(s)d\overline{N}_i(s),$$

where

$$\overline{K}_{y,b}(y-s) = \frac{a_2(y)-a_1(y)(y-s)}{a_0(y)a_2(y)-\{a_1(y)\}^2}K_b(y-s),$$

with

$$K_b(y-s) = \frac{1}{b}K(\frac{y-s}{b}), \tag{1}$$

and

$$a_j(y) = \sum_{i=1}^{n}\int K_b(y-s)(y-s)^j\overline{Y}_i(s)ds,$$

and

$$\widehat{S}(s) = \prod_{t \leq s} \left\{ 1 - d\widehat{\Lambda}(t) \right\},$$

being the Kaplan-Meier estimate of the survival function, with

$$\widehat{\Lambda}(s) = \sum_{i=1}^{n} \int_0^s \left\{ Y^{(n)}(t) \right\}^{-1} d\overline{N}_i(t),$$

where $Y^{(n)}(t) = \sum_{i=1}^{n} \overline{Y}_i(s)$.

2.3. Implementing with discrete data

In most real life applications we only have discretized versions of the stochastic processes Y_i and N_i available. First we need to define the relevant discretized time points $H_1, \ldots, H_K$ and the corresponding differences $h_k = H_k - H_{k-1}$ for $k \in \{1, \ldots, K\}$, with $H_0 = 0$. We define $H_K = \inf\left(t; F_{\widehat{\theta}}(t) = 1\right)$ for any plausible survival function F_θ.

Discretized data are often defined as occurrences and exposures. Let respectively

$$O_k = \sum_{i=1}^{n} \int_{H_{k-1}}^{H_k} dN_i(x)$$

and

$$E_k = \sum_{i=1}^{n} \int_{H_{k-1}}^{H_k} Y_i(x) dx.$$

Now assume that we only observe these discrete occurrences – the O_k's – and exposures – the E_k's. Then a natural approximation of the log likelihood function $l(\theta)$ above to our discrete observations would be

$$l_d(\theta) = \sum_k \left\{ \log \mu_\theta \left(H_k^* \right) \right\} O_k - \sum_k \mu_\theta \left(H_k^* \right) E_k,$$

where $H_k^* = \left(H_{k-1} + H_k \right) / 2$.

Now consider discretized time points on the axis transformed by $F_{\widehat{\theta}}$. Let $\overline{H}_k = F_{\widehat{\theta}}^*(H_k)$. $\overline{h}_k = \overline{H}_k - \overline{H}_{k-1}$ and $\overline{H}_k^* = \left(\overline{H}_{k-1} + \overline{H}_k \right) / 2$ for $k \in \{1, \ldots, K\}$. Note that $\overline{H}_K = 1$. Also note that often the discrete time points are equidistant before the time transformation but not thereafter.

On the transformed axis with time, the series $H_1^*, \ldots, H_K^*$ is transformed into $\overline{H}_1^*, \ldots, \overline{H}_K^*$. We will have occurrences

$$\overline{O}_k = O_k$$

and exposures

$$\overline{E}_k = E_k * \overline{h}_k / h_k.$$

Assume that we were given the true c.d.f. with very large risk exposures E_k. Then on the original axis $O_k \sim \mu_{\widehat{\theta}}(H_k^*) E_k h_k$ while on the transformed axis $\overline{O}_k = O_k \sim \alpha_{\widehat{\theta}}(\overline{H}_k^*) E_k \overline{h}_k$. If the model were the true one, the hazard rates $\overline{O}_k / \overline{E}_k$ on the transformed axis would be equal to $1 / (1 - \overline{H}_k^*)$, and hence the density functions would be constant at 1.

The local linear density estimator on the transformed axis will in the discrete case be defined as

$$\widehat{f}_d(y) = \sum_k \overline{K}_{d,y,b}(y - \overline{H}_k^*) \widehat{S}_d^t(\overline{H}_k^*) \overline{O}_k, \tag{2}$$

where

$$\overline{K}_{d,y,b}(y-s) = \frac{a_{2,d}(y) - a_{1,d}(y)(y-s)}{a_{0,d}(y) a_{2,d}(y) - \{a_{1,d}(y)\}^2} K_b(y-s),$$

$$a_{j,d}(y) = \sum_{k=1}^{K} K_b(y - \overline{H}_k^*)(y - \overline{H}_k^*)^j \overline{E}_k$$

and

$$\widehat{S}_d^t(\overline{H}_k^*) = 0.5 \left\{ \widehat{S}_d^t(\overline{H}_{k-1}) + \widehat{S}_d^t(\overline{H}_k) \right\} = 0.5 \left[\exp\left\{ -\sum_{i=1}^{k-1} \overline{h}_i \frac{\overline{O}_i}{\overline{E}_i} \right\} + \exp\left\{ -\sum_{i=1}^{k} \overline{h}_i \frac{\overline{O}_i}{\overline{E}_i} \right\} \right].$$

The choice of the bandwidth b depends on the availability of data. Large countries have a large risk exposure; then most of the deviation between the density estimate and 1 can be attributed to model uncertainty. In such cases, no or hardly any smoothing is required and b can be small. For not so densely populated countries with small risk exposure, on the other hand, proper smoothing – with a larger bandwidth – is needed to compensate for parameter uncertainty.

3. Mixed hazard models

In an individual frailty model the individual effect for a life's mortality acts multiplicatively. Assume that a cohort consists of n individuals. Then for the ith person of the cohort, the individual effect is represented by the random variable Z_i and the conditional force of mortality at age x, given $Z_i = z_i$, is given by

$$\mu(x, z_i) = z_i \mu(x), \quad i \in \{1, \ldots, n\},$$

with $\mu(x)$ denoting the standard force of mortality at age x – which is the force of mortality of a life with frailty level 1 – and all Z_i independent and identically distributed, with a mean equal to 1.

In this paper we will assume that the individual hazard is of the form

$$\mu(x) = \exp(a_0 + a_1 x + a_2 x^2). \tag{3}$$

In the notation of Forfar et al. (1988) this model is labelled GM(0,3). Note that the special case $a_2 = 0$ leads to the Gompertz model (GM(0,2)). The structure in (3) forms the basis for national and international mortality modelling in Jarner and Kryger (2011).

We have $d\mu(x)/dx = \mu(x)(a_1 + 2a_2 x)$. It is reasonable to assume that mortality is increasing as a function of age. This would imply $a_1 \geq 0$ and $a_2 \geq 0$. Nonnegative estimates of a_1 and a_2 are also obtained in Jarner and Kryger (2011). The relative change of mortality as a function of age x – defined in Horiuchi and Coale (1990) as $k(x) = d \ln \mu(x) / dx$ – is a linear function of age: $k(x) = a_1 + 2a_2 x$.

The cohort mortality at age x is given as $\mu_\theta(x) = E[Z|x] \cdot \mu(x)$, where $E[Z|x]$ denotes the mean frailty of lives surviving to age x. Let L_Z denote the Laplace transform of frailty at birth, i.e. $L_Z(s) = E[\exp(-sZ)]$. It can then be shown, see e.g. Hougaard (1984), that

$$E[Z|x] = -\frac{L_Z'[M(x)]}{L_Z[M(x)]}$$

with

$$M(x) = \int_0^x \mu(s)\, ds.$$

Hence the cohort mortality can be easily calculated for all frailty specifications with known Laplace transform. In the literature, the Gamma distribution has been by far the most popular specification in the frailty model. This is partly due to its mathematical tractability. Abbring and Van den Berg (2007) show that, under mild conditions

regarding regular variation, for several frailty distributions, Gamma is the limiting frailty distribution of survivors. Therefore, Gamma frailty is in some sense a natural choice. We will also be using Gamma frailty as one of the bases of our numerical illustrations in the next section.

Since the model (3) contains a scaling parameter, a_0, we will assume, without loss of generality, that the distribution of frailty at birth has mean 1 and variance σ^2. If frailty is Gamma, then $L_Z(s) = \left(1+\sigma^2 s\right)^{-1/\sigma^2}$, while $E[Z|x] = \left(1+\sigma^2 M(x)\right)^{-1}$.

Another common choice for frailty in the literature on survival models is the Inverse Gaussian distribution. This specification has also been discussed by Butt and Haberman (2004). Under the above assumption, $L_Z(s) = \exp\left[\sigma^{-2}\left(1-\sqrt{1+2\sigma^2 s}\right)\right]$ and $E[Z|x] = \left\{1+2\sigma^2 M(x)\right\}^{-1/2}$.

Disregarding heterogeneity implies $L_Z(s) = e^{-s}$ and $E[Z|x] \equiv 1$. This is a special case of the Gamma distribution, obtained by taking the limit $\sigma^2 \downarrow 0$. Obviously, this case of no frailty should give a worse fit than Gamma. In the numerical application we will show how this transpires by comparing the resulting local linear density estimators on the transformed scale.

4. Application to mortality data

We analyze the mortality data of four countries differing significantly in terms of population size, namely United States (US), United Kingdom (UK), Denmark and Iceland.

For each country, the data set consists of female period mortality data, obtained from the Human Mortality Database, and concerning the calendar year 2006. Since we are primarily interested in adult and old age mortality, only the ages from 40 to 110 are included. The exposed to risk at age 40 (defined before as E_1) are equal to $2,116,995.31$ (US), $478,424.36$ (UK), $43,463.83$ (Denmark) and $2,193.00$ (Iceland). So, roughly, the largest country is about 1,000 times as large as the smallest country.

For Gamma frailty, Inverse Gaussian frailty and no frailty, we estimate the parameters from maximum likelihood following Borgan (1984) and perform a visual inspection of its fit based on the transformation approach. The parameter estimates are shown in Table 1. As basic kernel function, we choose

$$K(x) = \frac{3003}{2048}\left(1-x^2\right)^6 I_{[-1,1]}(x). \tag{4}$$

For further details about this specific kernel, consult Nielsen et al. (2009). We have selected (4) for our example, but the actual choice of basic kernel function is not so important.

Table 1: *Estimates of parameters.*

Country/frailty model	a_0	a_1	a_2	σ^2
United States				
No frailty	−9.0389	0.04975	0.0003014	–
Gamma	−8.1373	0.02014	0.0005402	0.1193
Inverse Gaussian	−9.0388	0.04975	0.0003014	2.3966e-11
United Kingdom				
No frailty	−10.1631	0.06766	0.0002594	–
Gamma	−8.5070	0.01479	0.0006729	0.1632
Inverse Gaussian	−8.5047	0.01374	0.0006914	0.2488
Denmark				
No frailty	−10.1153	0.07338	0.0001991	–
Gamma	−9.7088	0.06034	0.0003021	0.04007
Inverse Gaussian	−9.7390	0.06125	0.0002954	0.04062
Iceland				
No frailty	−10.1777	0.05911	0.0003532	–
Gamma	−10.1136	0.05708	0.0003689	0.005556
Inverse Gaussian	−10.1777	0.05911	0.0003532	4.2043e-08

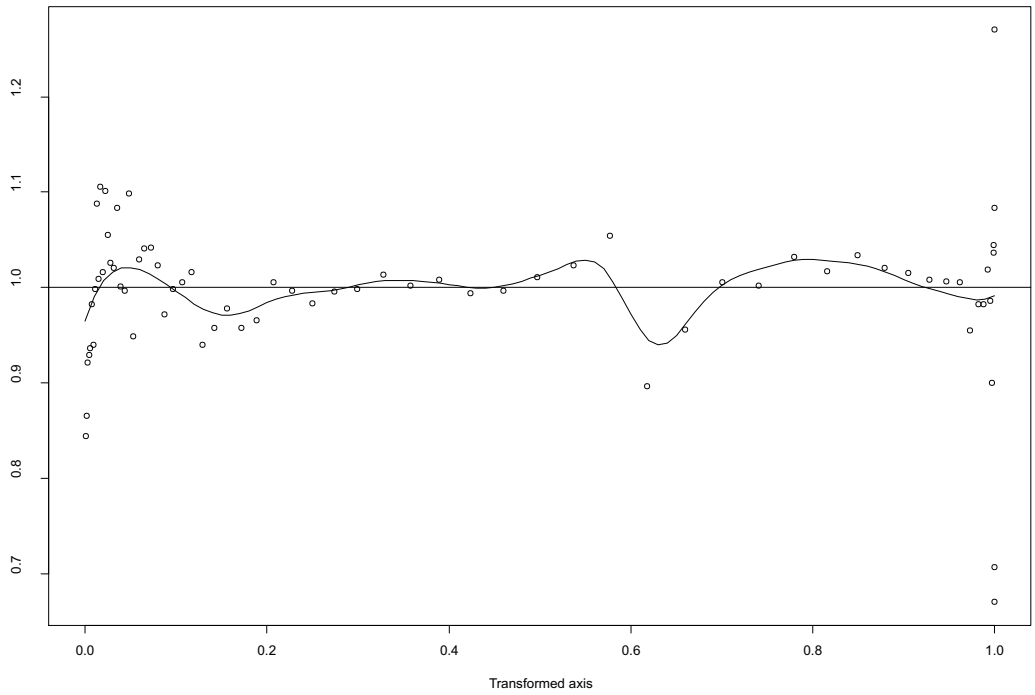

Figure 1: *United Kingdom: Values of density function obtained from the transformed observations (points) and local linear density estimator (curve) as in (2), with $b = 1/9$, all on the transformed scale, using Gamma frailty.*

Now we assume that Gamma frailty is suitable for the data at hand, and we want to test that assumption through our visual inspection technique. We use the UK as an illustrative example. Figure 1 displays the values of the density functions of the 71 transformed observations on the transformed scale, using Gamma frailty. For observation k, with $k \in \{1,\ldots,K\}$, the x-coordinate is equal to $\overline{H}_k^*$, while the corresponding y-coordinate is equal to $\left(1-\overline{H}_k^*\right) * \overline{O}_k / \overline{E}_k$.

As explained before, if the model assumed were the true one and risk exposure were infinite, these y-coordinates would all be equal to unity. Deviations from unity essentially arise from two sources: the extent to which the assumed/estimated model deviates from the true underlying model, and the noise in data caused by the stochastic nature of death. We are interested in assessing the first kind of (systematic) deviations, and for this purpose we want to reduce the second kind of (unsystematic) deviations.

Assume for a moment that all deviations could be taken at face value, i.e. that there were no unsystematic deviations. How would model deviations then manifest themselves on the transformed scale? Assume the model has density f with cdf F, while the true density is g. Then the density of transformed data is $\left(g/f\right)\left(F^{-1}(u)\right)$ for $0 < u < 1$. We are particularly interested in the behaviour in the right tail of the distribution. There are three possibilities:

- If the model overestimates the density of dying old then the density of transformed data will be below 1 in the right tail.
- If the model estimates the density of dying old correctly then the density of transformed data will be close to 1 in the right tail.
- If the model underestimates the density of dying old then the density of transformed data will be above 1 in the right tail.

Small risk exposure and consequently noisy data is a well known problem for very high ages, even for large countries like the US. This is why the observed density points start to deviate significantly from 1 when the x-coordinate approaches 1, bearing in mind that the probability of death before attaining a very advanced age is close to 1. The purpose of the density estimator is to reduce the noise in data to get a clearer picture of the performance of the model. The reason for working on the transformed scale is that the density can be more effectively estimated and with smaller bias on the unit interval than on the original scale.

4.1. Effect of bandwidth

A non-smoothed density estimate is obtained by simply connecting these points, but this will lead to an irregular pattern in all cases. The points are displayed together with the local linear density estimator based on kernel smoothing as in equation (2). In Figure 1, the value for the bandwidth b of $1/9$ has been selected by eye-ball, according to whatever looks best for this particular data set.

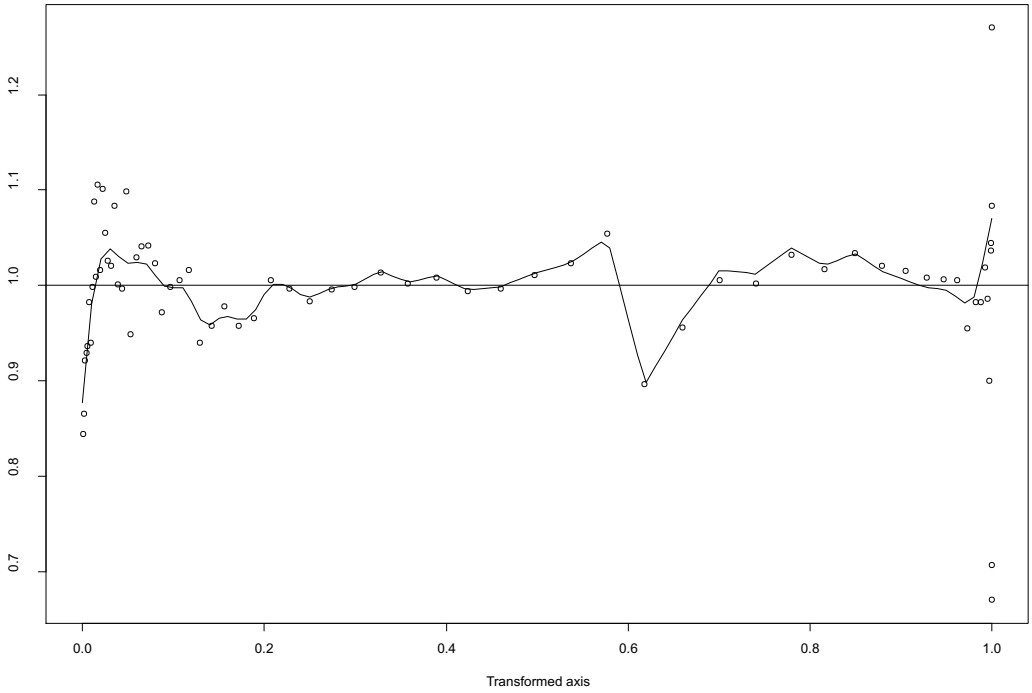

Figure 2: *United Kingdom: Values of density function obtained from the transformed observations (points) and local linear density estimator (curve) as in (2), with* $b = 1/24$*, all on the transformed scale, using Gamma frailty.*

The importance of choosing an appropriate value for the bandwidth is illustrated in the next two diagrams. Figure 2 shows the same points alongside the density estimator with smaller bandwidth $b = 1/24$. The lack of smoothness is evident: the estimator seems to be a set of line segments connecting several points.

Figure 3, on the other hand, gives the density estimator with larger bandwidth $b = 1/3$. This is an example of over-smoothing: features are displayed that do not reflect the characteristics of the data.

However, regardless of the chosen bandwidth the estimated density is close to unity in all three cases, and hence we would judge the model to provide a good description of old age mortality for the UK data.

The bandwidth choices for the other countries are $1/12$ (US), $1/6$ (Denmark) and $1/2$ (Iceland). Note that the bandwidths go up with decreasing population size for reasons as stated above. Observe that the set of observed density points for Iceland contains a lot of outliers, whence the choice of a large bandwidth.

While these bandwidths have been selected by eye-ball, we also reproduced automatic bandwidth selectors. The cross-validation procedure of Nielsen et al. (2009) breaks down and undersmooths way too much in our case. However, we adapted the Do-validation bandwidth selection procedure of Mammen et al. (2011) to our case. Except

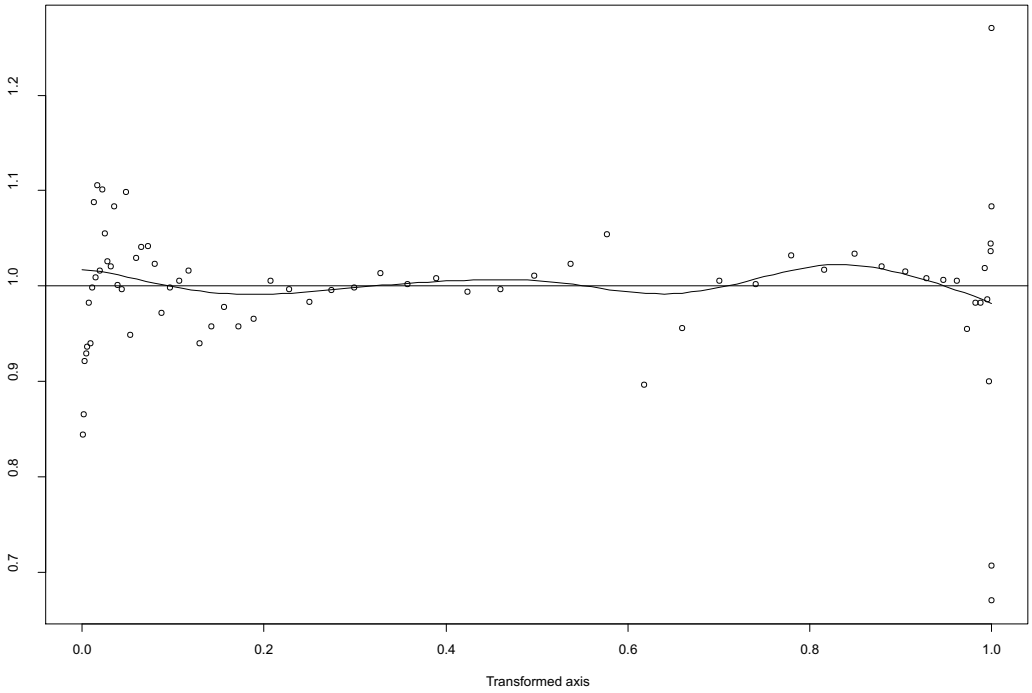

Figure 3: *United Kingdom: Values of density function obtained from the transformed observations (points) and local linear density estimator (curve) as in (2), with $b = 1/3$, all on the transformed scale, using Gamma frailty.*

for Iceland, we got almost exactly the same bandwidths as our eye-ball bandwidths. The automatic bandwidth selector for Iceland was somewhat higher than our eye-ball bandwidth selector. We did, however, in the end like our eye-ball selector for Iceland more and we therefore present that one here along with our other eye-ball selected bandwidths.

4.2. Old age mortality

In Figures 4 to 7, the local linear density estimate for Gamma frailty is compared with the one obtained in case of Inverse Gaussian and no frailty (with the same bandwidth, of course). Note that for the US the curves of Inverse Gaussian and no frailty practically coincide. Obviously this is due to the parameter estimates of a_0, a_1 and a_2 which are virtually the same for both models, while the estimate of the additional parameter σ^2 in Inverse Gaussian is very close to zero.

The overall impression regarding the four countries is best when it comes to the Gamma frailty survival model. In the USA case, the Gamma frailty parametric density is never more than two percent away from one on the transformed scale. This is not bad at all and most forecasters could accept deviances at this scale. Also, the USA data is quite abundant and we can allow ourself to work with a relatively small bandwidth.

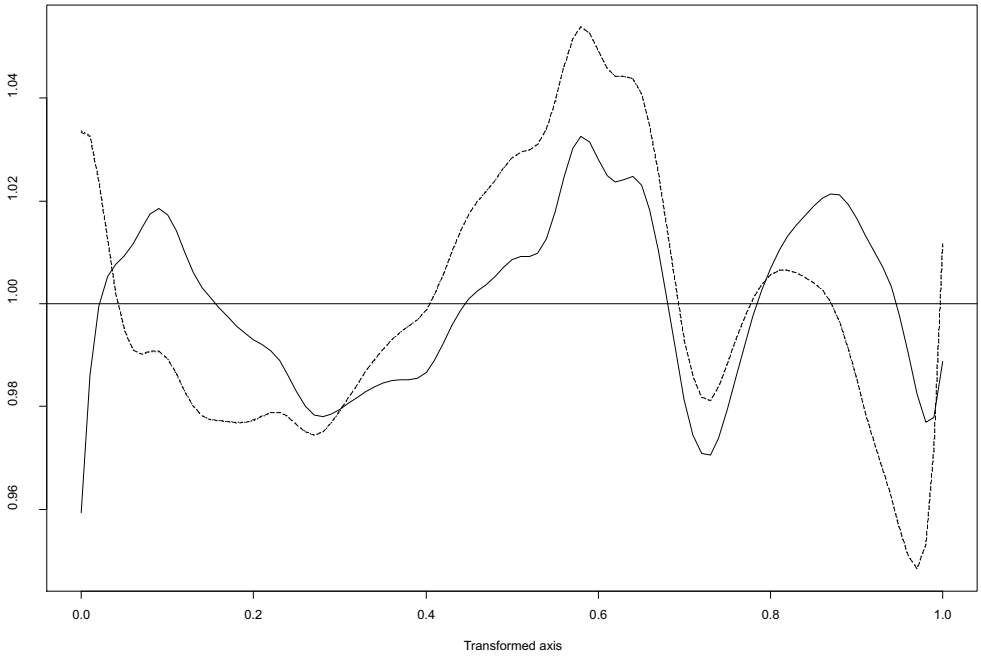

Figure 4: *United States: Local linear density estimator as in (2), with $b = 1/12$, on the transformed scale: Gamma frailty (solid) compared with Inverse Gaussian (dotted) and no frailty (dashed).*

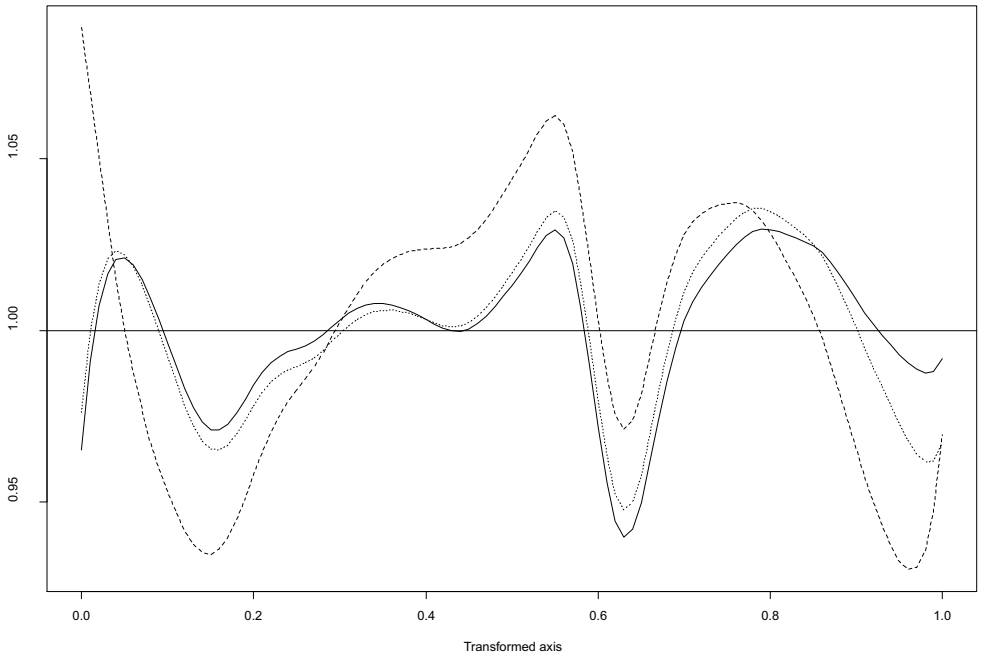

Figure 5: *United Kingdom: Local linear density estimator as in (2), with $b = 1/9$, on the transformed scale: Gamma frailty (solid) compared with Inverse Gaussian (dotted) and no frailty (dashed).*

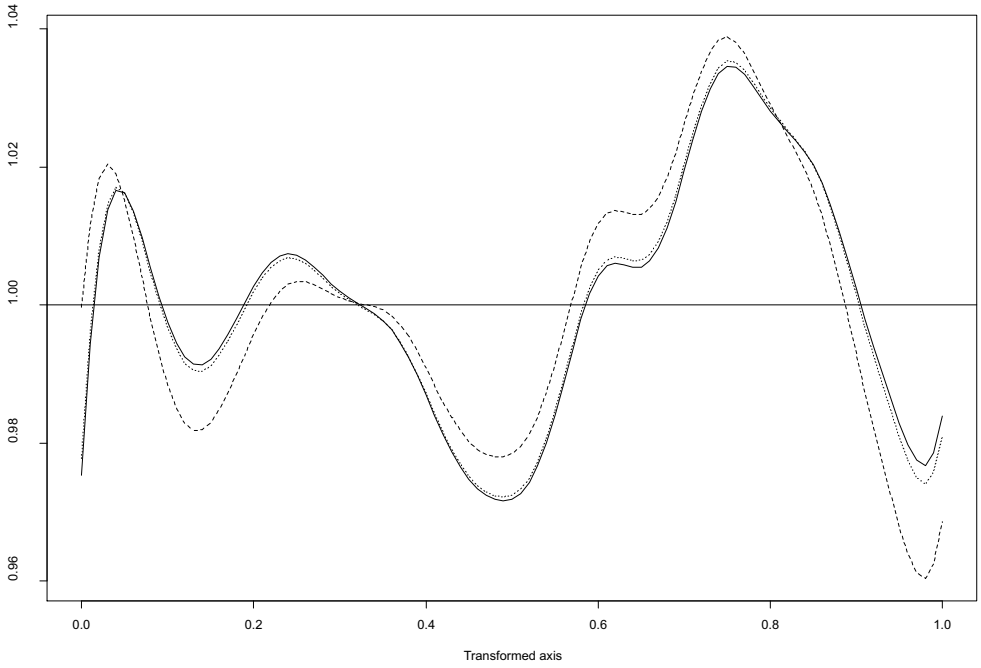

Figure 6: *Denmark: Local linear density estimator as in (2), with $b = 1/6$, on the transformed scale: Gamma frailty (solid) compared with Inverse Gaussian (dotted) and no frailty (dashed).*

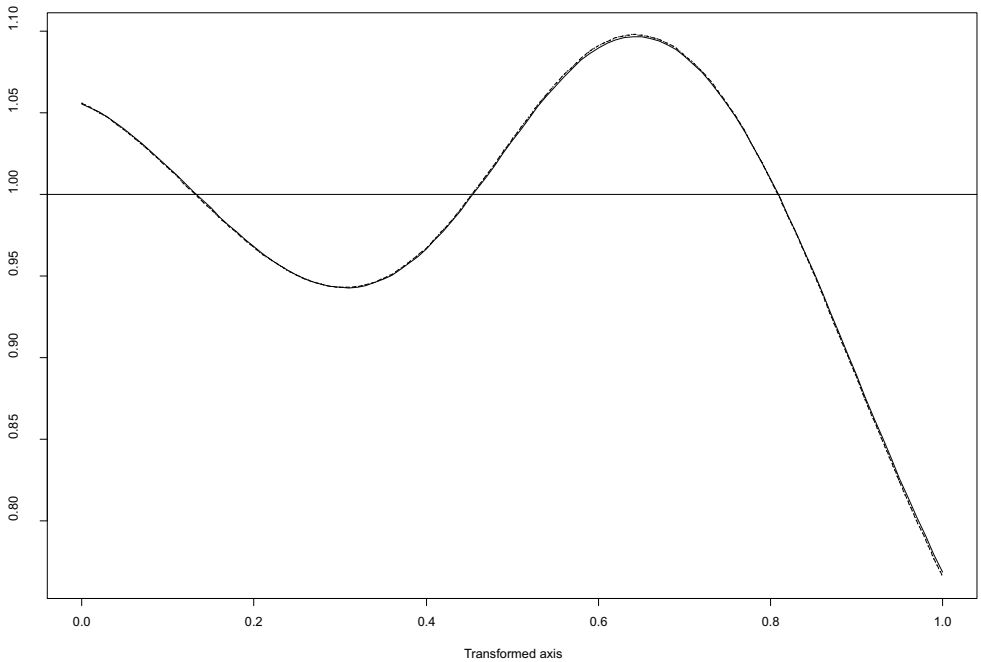

Figure 7: *Iceland: Local linear density estimator as in (2), with $b = 1/2$, on the transformed scale: Gamma frailty (solid) compared with Inverse Gaussian (dotted) and no frailty (dashed).*

Therefore, there is no reason to fear that we have smoothed too much and that should be the reason for the small deviance. This USA study gives us some confidence that the Gamma frailty survival model is working well also for smaller data sets, where bigger fluctuations are to be expected. For the United Kingdom the Gamma frailty survival model also fits relatively well, but now with deviances up to five percent. Surprisingly the Danish Gamma frailty survival density has very small deviances with the biggest being less than three percent. Iceland is another case, deviances up to 20% are found and the two frailty models do not seem to improve the fit compared to having no frailty at all. Overall the conclusion from the graphs is that the Gamma frailty makes the best fit, the Inverse Gaussian less so, but with both frailty models being superior to having no frailty at all. If we take a closer look at the tail of the three fitted Danish survival models at the transformed scale, we can get some further insight into the question posed in the introduction. It is indeed very clear that the flattening out of the Gamma frailty density in the tail helps the fit. The Gamma frailty version is much closer to one around the tail with about half the deviance from one compared to the no-frailty density version. In general, the performance of Inverse Gaussian is somewhat between that of Gamma and no frailty. For Iceland, the curves are almost identical, due to the small estimate of σ^2 and very similar estimates of the other parameters. In other words, for Iceland, the cases of Gamma frailty, Inverse Gaussian frailty and no frailty are nearly the same.

For US and UK the Gamma specification clearly provides the best description of data of the candidates considered. Both the Inverse Gaussian and no frailty alternative deviate substantially more in the right tail than Gamma frailty. These two specifications both overestimate old age mortality substantially, while the Gamma frailty seems to capture the old age mortality plateau evident in data. Moreover, the right tail deviations of Gamma frailty is of the same magnitude as deviations for younger age segments, while the right tail deviations for Inverse Gaussian and no frailty seems to diverge. While the picture is less clear for Denmark, the frailty densities also here improve the description of old age mortality. It also seems that without frailty the deviation diverges in the right tail, but the magnitude of deviation is much smaller than for US and UK. In contrast to US and UK, the Gamma and Inverse Gaussian essentially perform equally well. Thus we conclude that there is enough information in data to indicate the presence of heterogeneity, but not enough information to distinguish between the different kinds of heterogeneity.

Lastly, Iceland has so little exposure and so much uncertainty in data that even with the method derived in this paper we cannot distinguish between the models.

A critical part of the study concerns the performance of the estimator for advanced ages. To this end, for each country we calculate the second largest and largest points of intersection of the estimator with the horizontal line (i.e. the two largest roots of the equation $\widehat{f}_d(y) = 1$) and translate this back into the corresponding ages. For comparability, we have left out in this investigation the late spike of no frailty and the Inverse Gaussian in the USA case. The results are given in Table 2 below. It is quite clear that Gamma frailty densities in all case are having the last crossing point. This indicates

that the Gamma frailty density provides the best description of old age mortality among the considered models. In the USA case, we get almost to the age of 100 before our transformed density drops below one. The lowest last crossing for the gamma frailty is still quite high, namely 93 years. Above this last crossing point on the transformed scale, all the fitted parametric models seem to have too low densities. Thus, above the last crossing point our parametric models are overstating the possibility of dying. In other words, above the last crossing point all models seem to be on the safe side. In particular the Gamma frailty seems well behaved for annuity purposes. The density of very old are a bit too high, but rarely more than two percent, and these two percent are on the safe side when calculating for example annuities. Most of the extra old age mass is taken from the interval between the next last crossing point and the last crossing point, where the underlying parametric densities are overestimated in all cases. Therefore, while none of the densities are making a perfect fit, the Gamma frailty density is very close and with good properties for the annuity forecaster. It is on the safe side for the very old ages, with an overall annuity that seems to be close to the truth, overestimating the density in the very old ages, but compensating for that overestimation in an interval leading up to those old ages. Without frailty the deviations for old ages are substantially larger than with (gamma) frailty. The transformed density is below 1 which indicates that the probability of dying old is overestimated. At first glance this appears to be at odds with the fact that without frailty the old-age hazard is overestimated, cf. Figure 1. The explanation is that while the old-age hazard is overestimated the hazard is underestimated in the age groups below and therefore too many attain the (high) age of 90, say, after which they die too quickly. The model without frailty is on the safe when setting aside reserves for annuities for 40 year-olds, but if we were to use the same model for older age groups it would only be conservative up to a certain point. This clearly is not a desirable feature, and it illustrates the point that overstating the probability of dying old for one cohort is not necessarily a conservative assumption for other cohorts.

Notice that it would be hard to get this kind of detailed information from testing the underlying densities or even from graphical visualization techniques on the original scale. Therefore, our simple transformation technique has enabled us to comfort the statistician forecasting mortality models based on simple underlying parametric survival distributions.

Table 2: *Second largest and largest crossing point of density estimator with horizontal line at 1.*

Country	**Gamma**		**Inverse Gaussian**		**No frailty**	
	Second largest crossing point	Largest crossing point	Second largest crossing point	Largest crossing point	Second largest crossing point	Largest crossing point
US	92.37	98.68	92.14	95.21	92.14	95.21
UK	89.51	96.85	89.15	95.34	87.61	92.10
Denmark	85.53	94.77	85.47	94.67	84.91	93.63
Iceland	84.38	93.02	84.34	93.01	84.34	93.01

5. Conclusions

We have developed a new visual inspection technique of survival models. It generalizes developments of transformation techniques of i.i.d. data, see for example Bolancé et al. (2008, 2012a, 2012b, 2013). The method seems useful in many versions of follow-up studies, see for example Guillén et al. (2012) and Pinquet et al. (2011). We imagine it to be useful when the applied statistician wants the data to guide his intuition. The working methodology could be through running the knowledge loop cycle: Data→Visualization→New Assumption a number of times until the final assumptions seem intuitively reasonable and well behaved also according to more standard statistical techniques.

All the mortality projection models discussed in the Introduction involve both an age and a time dimension. As mentioned in the Introduction one can use our one-dimensional visualization technique for the age effect after having adjusted for the time effect and vice versa when visualizing the time effect. A full multidimensional version of our methodology is also possible. One could use multidimensional density estimation of filtered data to introduce a similar visual inspection technique to assessing the quality of mortality depending on both age and time. See for example Buch-Kromann and Nielsen (2012) for a recent multivariate density estimator that could be used in our visual diagnostic step after having transformed our data with our favourite forecasting mortality model.

Transformations and visual fitting as developed in this paper would also seem relevant in other areas of actuarial science as, for example, reserving, see the recent papers Martínez-Miranda et al. (2012) and Kuang et al. (2011).

Acknowledgement

This project was funded by a research grant from The Actuarial Foundation and the Society of Actuaries.

References

Abbring, J. H. and Van den Berg, G. J. (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika*, 94 (1), 87–99.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Bolancé, C., Guillen, M. and Nielsen, J. P. (2008). Inverse beta transformation in kernel density estimation. *Statistics and Probability Letters*, 78, 1757–1764.

Bolancé, C., Guillén, M., Nielsen, J. P. and Gustafsson, J. (2012a). *Quantitative Operational Risk Models*. Chapman and Hall/CRC Finance Series, New York.

Bolancé, C., Ayuso, M. and Guillén, M. (2012b). A nonparametric approach to analysing operational risk with an application to insurance fraud. *The Journal of Operational Risk*, 7 (1), 1–16.

Bolancé, C., Guillén, M., Gustafsson, J. and Nielsen, J. P. (2013). Adding prior knowledge to quantitative operational risk models. *The Journal of Operational Risk*, 8 (1), 17–32.

Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, 11, 1–16.

Buch-Kromann T. and Nielsen, J. P. (2012). Multivariate density estimation using dimension reducing information and tail flattening transformations for truncated and censored data. *Annals of the Institute of Mathematical Statistics*, 48 (1), 167–192.

Butt, Z. and Haberman, S. (2004). Application of frailty-based mortality models using Generalized Linear Models. *ASTIN Bulletin*, 34 (1), 175–197.

Cairns, A. J. G., Blake, D. and Dowd, K. (2006). A two factor model for stochastic mortality and parameter uncertainty: theory and calibration. *The Journal of Risk and Insurance*, 73 (4), 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13 (1), 1–35.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2011). Mortality density forecasts: an analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48, 355–367.

Currie, I. D. (2006). Smoothing and forecasting mortality rates with P-splines. *Talk given at the Institute of Actuaries*, June 2006. http://www.actuaries.org.uk

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2010a). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, 47, 255–265.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2010b). Backtesting stochastic mortality models: an ex post evaluation of multi-period ahead density forecasts. *North American Actuarial Journal*, 14 (3), 281–298.

Forfar, D. O., McCutcheon, J. J. and Wilkie, A. D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115, 1–149.

Gámiz-Pérez, M. L., Martínez-Miranda, M. D. and Nielsen, J. P. (2013a). Smoothing survival densities in practice. *Computational Statistics and Data Analysis*, 58 (1), 368–382.

Gámiz Pérez, M. L, Mammen, E., Martínez Miranda, M. D. and Nielsen, J. P. (2013b). Do-validating local linear hazards. *Submitted preprint.*

Gámiz-Pérez, M. L., Janys, L., Martínez-Miranda, M. D. and Nielsen, J. P. (2013c). Smooth marker dependent hazard estimation in praxis. *Computational Statistics and Data Analysis*, *Forthcoming.*

Guillén, M., Nielsen, J. P., Scheike, T. H. and Pérez-Marín, A. M. (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Systems with Applications*, 39 (3), 3551–3558.

Haberman, S. and Renshaw, A. E. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48, 35–55.

Horiuchi, S. and Coale, A. J. (1990). Age patterns of mortality for older women: an analysis using the age-specific rate of mortality change with age. *Mathematical Population Studies*, 2 (4), 245–267.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, 71 (1), 75–83.

Jarner, S. F. and Kryger, E. M. (2011). Modelling adult mortality in small populations: The SAINT model. *ASTIN Bulletin*, 41 (2), 377–418.

Jones, B. L. (1998). A model for analyzing the impact of selective lapsation on mortality. *North American Actuarial Journal*, 2 (1), 79–86.

Kuang D., Nielsen, B. and Nielsen, J. P. (2011). Forecasting in an extended chain-ladder-type model. *Journal of Risk and Insurance*, 78 (2), 345–359.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87 (419), 659–671.

Li, J. S.-H., Hardy, M. R. and Tan, K. S. (2009). Uncertainty in mortality forecasting: an extension of the Lee-Carter approach. *ASTIN Bulletin*, 39 (1), 137–164.

Mammen, E., Martínez-Miranda, M. D., Nielsen, J. P. and Sperlich, S. (2011). Do-validation for kernel density estimation. *Journal of the American Statistical Association*, 106 (494), 651–660.

Martínez-Miranda, M. D., Nielsen, J. P. and Wüthrich, M. V. (2012). Statistical modelling and forecasting of outstanding liabilities in non-life insurance. *SORT-Statistics and Operations Research Transactions*, 36 (2), 195–218.

Nielsen, J. P., Tanggaard, C. and Jones, M. C. (2009). Local linear density estimation for filtered survival data. *Statistics*, 43 (2), 167–186.

Olivieri, A. (2006). Heterogeneity in survival models-applications to pensions and life annuities. *Belgian Actuarial Bulletin*, 6 (1), 23–39.

Pinquet, J., Guillén, M. and Ayuso, M. (2011). Commitment and lapse behavior in long-term insurance: a case study. *Journal of Risk and Insurance*, 78 (4), 983–1002.

Plat, R. (2009). On stochastic mortality modelling. *Insurance: Mathematics and Economics*, 45, 393–404.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38, 556–570.

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439–454.

Wang, S. and Brown, R. L. (1998). A frailty model for projection of human mortality improvements. *Journal of Actuarial Practice*, 6, 221–241.